COMP5313 — Lab 4

Web Pages Popularity

The goal of this tutorial is to understand what impact the ranking of Web pages based on hub and authority update rules.

Exercise 1: Computing authority scores

Show the values that you get if you run two rounds of computing hub and authority values on the network of Web pages in Figure 1. (That is, the values computed by the k-step hub-authority computation when we choose the number of steps k to be 2.)
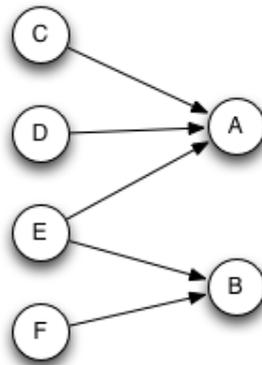

Figure 1: A simplified example of a Web graph

Show the values both before and after the final normalization step, in which we divide each authority score by the sum of all authority scores, and divide each hub score by the sum of all hub scores. (We will call the scores obtained after this dividing-down step the normalized scores. It's fine to write the normalized scores as fractions rather than decimals.)

Duration: 20 min

Exercise 2: Creating Web pages

Now we come to the issue of creating pages so as to achieve large authority scores, given an existing hyperlink structure.

In particular, suppose you wanted to create a new Web page X, and add it to the network in Figure 1, so that it could achieve a (normalized) authority score that is as large as possible. One thing you might try is to create a second page Y as well, so that Y links to X and thus confers authority on it. In doing this, it is natural to wonder whether it helps or hurts X's authority to have Y link to other nodes as well. Specifically, suppose you add X and Y to the network in Figure 1. In order to add X and Y to this network, one needs to specify what links they will have. Here are two options; in the first option, Y links only to X, while in the second option, Y links to other strong authorities in addition to X.

1. Add new nodes X and Y to Figure 1; create a single link from Y to X; create no links out of X.

2. Add new nodes X and Y to Figure 1; create links from Y to each of A, B, and X; create no links out of X.

For each of these two options, we would like to know how X fares in terms of its authority score. So, for each option, show the normalized authority values that each of A, B, and X get when you run the 2-step hub-authority computation on the resulting network (as in the former question). (That is, you should perform the normalization step where you divide each authority value down by the total.)

Duration: 15 min

Exercise 3: Creating hyperlinks

Suppose instead of creating two pages, you create three pages X, Y, and Z. Describe a strategy for adding three nodes X, Y, and Z to the network in Figure 1, with choices of links out of each, so that when you run the 2-step hub-authority computation (as in former questions), and then rank all pages by their authority score, node X shows up in second place.

Is there another strategy of choosing X, Y and Z outgoing edges so that X shows up in first place?

Duration: 15 min